M E M O R A N D U M

TO: **The Kentucky Board of Education**

FROM: **Gene Wilhoit**

DATE: **August 30, 2006**

SUBJECT: **Reliability Coefficients for the Kentucky Core Content Test and Other Major Standardized Tests**

Comparison of KCCT Reliability at the Student Level to Reliability of Commercial Tests

In the packet of materials you received for the September 6 KBE meeting, dated August 24, the staff note titled "Revisions to the Accountability System" had four attachments. The fourth one, Attachment D, titled "Validity and Reliability", addressed the questions of what each term means and contained a chart on page 3 presenting the reliability coefficients at the student level for the Kentucky Core Content Test by grade and subject area from 2001 through 2005.

To give you some comparative information, we have attached the same chart showing the reliability coefficients at the student level for the KCCT and additional charts presenting the reliability coefficients at the student level for some major standardized tests as follows: ACT, SAT, the California Achievement Test/5 (CAT/5), and CTB *TeraNova*. These charts came from the publishers of these tests.

Most commercial standardized tests report reliability coefficients at the student level that generally range between .80 and .90. As stated in the KCCT Technical Manual for 2002 and 2004, ". . . KCCT student-level reliability coefficients compare favorably with the reliability reported by the publishers of nationally norm-referenced tests . . .". See the attached charts for these comparisons.

Use of Multiple Sources of Evidence

Valid decisions about students should never be based on one test score alone, but should consider other evidence, such as additional test scores, student work, and teacher judgment. This point has been echoed on multiple occasions by the National Technical Panel on Assessment and Accountability (NTAPAA), especially for high stakes assessment. More information, particularly information from different sources, strengthens validity. Even when multiple sources of evidence are used in making high-stakes decisions about students,

e.g., passing to the next grade, participating in a course, getting a job, and so on, any test score used in making the decision, should be highly reliable.

Multiple Item Formats

The Kentucky Core Content Tests use multiple-choice and open-response items. Both types of items may test high-level thinking skills, but a number of advantages are associated with open-response items. They require students to construct their responses by tapping multiple facets of knowledge and to demonstrate skills appropriate to the discipline. Many, though perhaps not all, scholars feel that this feature strengthens the validity of test-based inferences by allowing fuller representation of the student's knowledge and skills. In addition, open-response items require students to demonstrate that they can communicate, consistent with provisions of the Kentucky Academic Expectations.

Because multiple-choice items have only one correct answer, no judgment is required to score them. The use of multiple-choice items is, therefore, associated with higher reliability. Student responses to open-response items, on the other hand, are scored by trained raters who use professionally designed rubrics. Raters must use personal judgment in assigning scores to students' written responses. The use of personal judgment in scoring means that there is more room for variability among raters, which is viewed as possible error. However, with comprehensive training and common scoring practices, reliability is increased. The reliability of tests using open-response items may 'take a hit,' however small, reducing their reliability coefficients slightly as compared to multiple-choice-only tests.

If you have questions on this addendum to the validity and reliability materials, please let us know.

I look forward to our discussion of accountability revisions at the September 6 meeting.

Kentucky Core Content Student-Level Reliability Coefficients

Table 14-2
KCCT Reliability 2001 – 2005
Median and Range Coefficient Alpha Across Forms* by Grade and Subject

Core Content Test by Grade		2001		2002		2003		2004		2005	
		Median[1]	Range	Median[1]	Range	Median[1]	Range	Median[1]	Range	Median[1]	Range
4/5	**Reading**	.88	.87-.88	.88	.86-.88	.86	.85-.87	.87	.84-.87	.86	.85-.89
	Mathematics	.87	.86-.88	.87	.86-.88	.87	.86-.87	.86	.83-.88	.87	.84-.88
	Science	.84	.80-.85	.83	.81-.84	.83	.82-.85	.84	.81-.85	.82	.81-.83
	Social Studies	.84	.84-.85	.85	.83-.86	.84	.83-.85	.83	.82-.84	.83	.81-.84
	Arts & Hum	.66	.63-.67	.66	.63-.71	.66	.62-68	.65	.61-.73	.64	.60-.67
	PL/VS	.63	.53-.67	.69	.67-.73	.61	.50-.64	.59	.49-.67	.58	.53-.62
7/8	**Reading**	.87	.87-.88	.87	.87-.88	.86	.85-.87	.86	.85-.87	.86	.85-.86
	Math	.89	.88-.90	.89	.88-.90	.89	.88-.89	.89	.88-.90	.89	.88-.90
	Science	.84	.83-.86	.86	.84-.86	.85	.84-.86	.85	.84-.86	.85	.83-.87
	Social Studies	.89	.87-.89	.88	.87-.89	.88	.87-.89	.88	.86-.88	.87	.85-.88
	Arts & Hum	.70	.66-.73	.69	.67-.73	.67	.59-.73	.66	.61-.73	.68	.63-.72
	PL/VS	.70	.66-.74	.71	.67-.74	.68	.63-.73	.67	.62-.71	.66	.63-.72
10/12	**Reading**	.88	.87-.89	.88	.88-.89	.87	.87-.88	.89	.88-.91	.89	.87-.89
	Mathematics	.88	.85-.89	.89	.87-.89	.89	.88-.89	.89	.88-.90	.88	.88-.90
	Science	.84	.82-.85	.85	.81-.86	.84	.82-.85	.84	.82-.84	.84	.84-.85
	Social Studies	.88	.87-.88	.89	.88-.89	.88	.87-.89	.88	.87-.89	.88	.87-.89
	Arts & Hum	.67	.61-.72	.69	.65-.72	.66	.62-.68	.67	.57-.71	.66	.60-.72
	PL/VS	.64	.60-.68	.65	.62-.68	.64	.56-.67	.63	.57-.71	.62	.51-.65

[1]Median coefficient alpha is based upon operational matrix OR and MC items across test forms.
Six test forms are used in reading, mathematics, science and social studies; twelve forms in Arts and Humanities and Practical Living/Vocational Studies.

Chart is from the Kentucky Core Content Tests 2005 Technical Appendices.

ACT Student-Level Reliability Coefficients

Table 11-2
ACT Scale Score Reliability[1]

Test/Subtest	Scale Score Reliability		
	Median	Minimum	Maximum
English	.91	.90	.91
Usage/Mechanics	.84	.82	.84
Rhetorical Skills	.85	.85	.86
Mathematics	.91	.89	.91
Pre-Algebra/Elementary Algebra	.82	.78	.83
Intermediate Algebra/Coordinate Geometry	.71	.64	.74
Plane Geometry/Trigonometry	.74	.71	.78
Reading	.86	.86	.87
Social Studies/Sciences	.77	.73	.78
Arts/Literature	.78	.76	.80
Science Reasoning	.84	.82	.86

[1] ACT (1997). Assessment Technical Manual: Iowa City, IA: Author.

SAT I and SAT II Student-Level Reliability Coefficients

Table 14 - 4
SAT I & SATII Test Reliabilities[2]

Test	Number of Questions	Reliability Coefficient
SAT I		
Verbal	78	.91-.93
Math	60	.91-.93
SAT II		
Writing		.86 -.90
Multiple Choice	60	.88-.89
Essay	1	.77-.82
Literature	60-62	.87-.90
U.S. History	89-90	.91-.93
World History	94-95	.91-.94
Math Level IC	50	.88-.90
Math Level IIC	50	.87-.89
Biology		
Biology EB	79-80	.89-.91
Biology MB	80	.90-.91
Chemistry	85	.93-.94
Physics	74-75	.90-.92

[2] The College Board Data Tables
http://www.collegeboard.com/counselors/hs/sat/scorereport/scoredata.html
as reflected in the KCCT 2004 Technical Report

CAT/5 Survey A Student-Level Reliability Coefficients

Table 14 - 3
CAT/5 Survey A (Spring) Reliability[3]

Grade	Section	# of Items	Reliability
3	Reading	40	.87
	Vocabulary	20	.78
	Comprehension	20	.80
	Language	40	.85
	Language Mechanics	20	.78
	Language Expression	20	.74
	Mathematics	40	.84
	Math Computation	20	.73
	Math Concepts & Application	20	.76
	Science	20	.73
	Social Studies	20	.80
	Mathematics	40	.85
	Math Computation	20	.78
	Math Concepts & Application	20	.74
	Science	20	.79
	Social Studies	20	.74
6	Reading	40	.88
	Vocabulary	20	.79
	Comprehension	20	.81
	Language	40	.85
	Language Mechanics	20	.77
	Language Expression	20	.78
	Mathematics	40	.87
	Math Computation	20	.82
	Math Concepts & Application	20	.76
	Science	20	.74
	Social Studies	20	.84
9	Reading	40	.87
	Vocabulary	20	.79
	Comprehension	20	.80
	Language	40	.86
	Language Mechanics	20	.77
	Language Expression	20	.79
	Mathematics	40	.87
	Math Computation	20	.83
	Math Concepts & Application	20	.75
	Science	20	.69
	Social Studies	20	.81

[3] CTB Macmillan/McGraw-Hill. (1992). CAT/5 Technical Bulletin 1:Monterey, CA: Author.

***TeraNova* Performance Assessments Student-Level Reliability Coefficients**

Tera Nova Performance Assessments Reliability[4]				
Grade	**Level**	**Test**	**Average p-value**	**Reliability**
3	13/14	Reading	.60	.84
		Writing	.52	.89
		Mathematics	.38	.94
4	14/15	Reading	.49	.91
		Writing	.55	.91
		Mathematics	.30	.88
5	14/15	Reading	.54	.90
		Writing	.63	.91
		Mathematics	.40	.90
6	16/17	Reading	.52	.93
		Writing	.62	.90
		Mathematics	.37	.90
7	17/18	Reading	.58	.85
		Writing	.64	.93
		Mathematics	.40	.90
8	17/18	Reading	.61	.85
		Writing	.68	.92
		Mathematics	.47	.90
9	19/20	Reading	.52	.94
		Writing	.57	.91
		Mathematics	.37	.91
10	19/20	Reading	.55	.94
		Writing	.60	.92
		Mathematics	.41	.92
11	21/22	Reading	.61	.95
		Writing	.68	.75
		Mathematics	.41	.92
12	21/22	Reading	.64	.95
		Writing	.70	.74
		Mathematics	.44	.92

[4] CTB Macmillan/McGraw-Hill. (2001). Technical Report Performance Assessments *TerraNova*: Monterey, CA: Author.